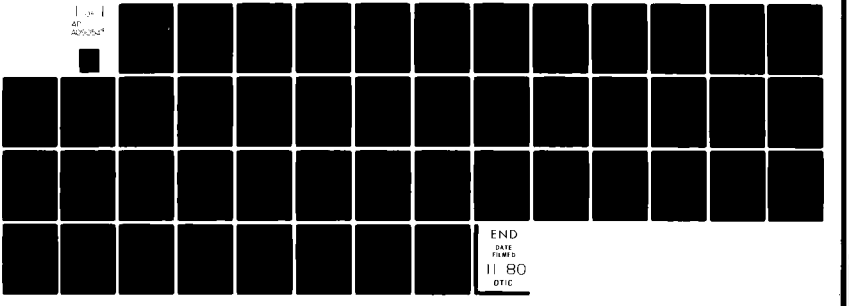
AD-A090 547 BROWN UNIV PROVIDENCE RI LEFSCHETZ CENTER FOR DYNAMI--ETC F/G 6/16
PARAMETER IDENTIFICATION TECHNIQUES FOR PHYSIOLOGICAL CONTROL S--ETC(U)
JUL 80 H T BANKS DAAG29-79-C-0161
UNCLASSIFIED AFOSR-TR-80-1001 NL



END
DATE FILMED
11 80
DTIC

(3)

PARAMETER IDENTIFICATION TECHNIQUES

for

PHYSIOLOGICAL CONTROL SYSTEMS*

LEVEL II

by

H. T. Banks+
Lefschetz Center for Dynamical Systems,
Division of Applied Mathematics
Brown University
Providence, R. I. 02912

July, 1980

---

* Lectures given at the American Mathematical Society Summer Seminar on Mathematical Aspects of Physiology, University of Utah, June 15-27, 1980.

+ Some of the research of the author reported herein was supported in part by the Air Force Office of Scientific Research under contract #[redacted]-AFOSR 76-3092[redacted], in part by the U.S. Army Research Office under contract #ARO-DAAG29-79-C-0161 and in part by the National Science Foundation under contract #NSF-MCS 7905774.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) UNCLASSIFIED

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR-80-1001 2. GOVT ACCESSION NO AD-A090 547	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PARAMETER IDENTIFICATION TECHNIQUES FOR PHYSIOLOGICAL CONTROL SYSTEMS.	5. TYPE OF REPORT & PERIOD COVERED Interim rept. 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) H.T. BANKS	8. CONTRACT OR GRANT NUMBER(s) 15 AFOSR-76-3092 DAAG29-77-C-0161
9. PERFORMING ORGANIZATION NAME AND ADDRESS DIVISION OF APPLIED MATHEMATICS BROWN UNIVERSITY PROVIDENCE, RHODE ISLAND 02912	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 16 17 61102F 2304 A1
11. CONTROLLING OFFICE NAME AND ADDRESS AIR FORCE OFFICE OF SCIENTIFIC RESEARCH BOLLING AIR FORCE BASE WASHINGTON, D.C.	12. REPORT DATE 11 Jul 80 (12) 47 13. NUMBER OF PAGES 44
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Parameter identification techniques for problems involving functional and partial differential equation models are discussed. The methods presented combine standard ordinary differential equation algorithms with Ritz-Galerkin ideas in reducing problems

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 401 [illegible]

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

for infinite dimensional state systems to finite dimensional state problems.

PARAMETER IDENTIFICATION TECHNIQUES

for

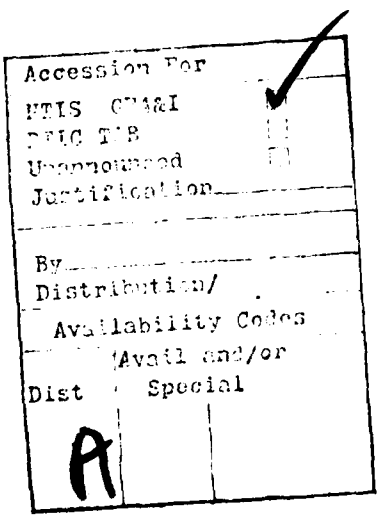
PHYSIOLOGICAL CONTROL SYSTEMS

by

H. T. Banks

ABSTRACT

Parameter identification techniques for problems involving functional and partial differential equation models are discussed. The methods presented combine standard ordinary differential equation algorithms with Ritz-Galerkin ideas in reducing problems for infinite dimensional state systems to finite dimensional state problems.



I. Introduction

In these lectures we discuss parameter identification problems for control systems that frequently arise in certain physiological models. By a parameter identification problem we shall mean that we are given a model with unknown parameters along with observations of the system that the model is supposed to represent and we must use these observations to determine values for the unknown parameters. Specifically, we shall focus our attention on problems for nonlinear models involving either (1) delay-differential and, more generally, functional differential equations (FDE), or (2) distributed parameter systems realized by partial differential equations (PDE). In the first case, we shall discuss techniques that are applicable to problems that contain unknown delays as well as unknown transport coefficients among the parameters. In the case of (2) we present methods for problems which contain unknown diffusion, solubility, and other transport coefficients. In both cases the methods can be used in problems with unknown parameters in the initial/boundary data.

Briefly, our presentation will be as follows. First we give motivating examples consisting of models arising in respiratory-circulatory physiology and renal physiology. Before discussing some inherent difficulties in parameter estimation for the resulting classes of problems, we review standard methods available for parameter identification of ordinary differential equation models. Included for mention in a least-squares formulation are descent

methods such as gradient, conjugate-gradient, and certain quasi-Newton algorithms. Finally, we explain how one can use Ritz-Galerkin ideas to employ the standard techniques in a way that circumvents the difficulties mentioned and produces convergent identification algorithms for problems involving nonlinear FDE and PDE.

We provide an adequate but not exhaustive bibliography. Many of the publications we have selected to cite here include rather extensive reference to the current research literature.

2. Models from physiology involving delay and distributed parameter systems

We first consider respiratory models for humans which are formulated with the overall schematic depicted in Figure 1 in mind. These models are often ones which combine (1) the respiratory system (based on gas dynamics) with compartments such as alveolar space, dead space, airways, and atmosphere which are connected to a "gas pump" and (2) the circulatory system (often based on transport theory and fluid dynamics) with compartments such as pulmonary vein, left heart, aorta, arteries, tissue, veins, right heart, and pulmonary artery. The respiratory and circulatory components are connected via a diffusion compartment representing exchange of the principal gases N_2,O_2, H_2O and CO_2 which are the subject of investigations in these models [13], [17], [19], [24], [25],[26],[27],[33],[34]. This gas exchange takes place, of course, during the flow of the blood from the pulmonary artery to the pulmonary vein as it perfuses the alveoli.

Many of the models entail systems of ordinary differential equations and the more realistic ones usually take into account transport delays and thus involve delay-differential equations or functional differential equations. Briefly one uses understanding of the basic biomechanical and biochemical mechanisms of the systems to write mass balance equations. These equations relate the alveolar partial pressures $P_{A(j)}$, brain concentrations $C_{B(j)}$, and tissue concentrations $C_{T(j)}$ for the various species

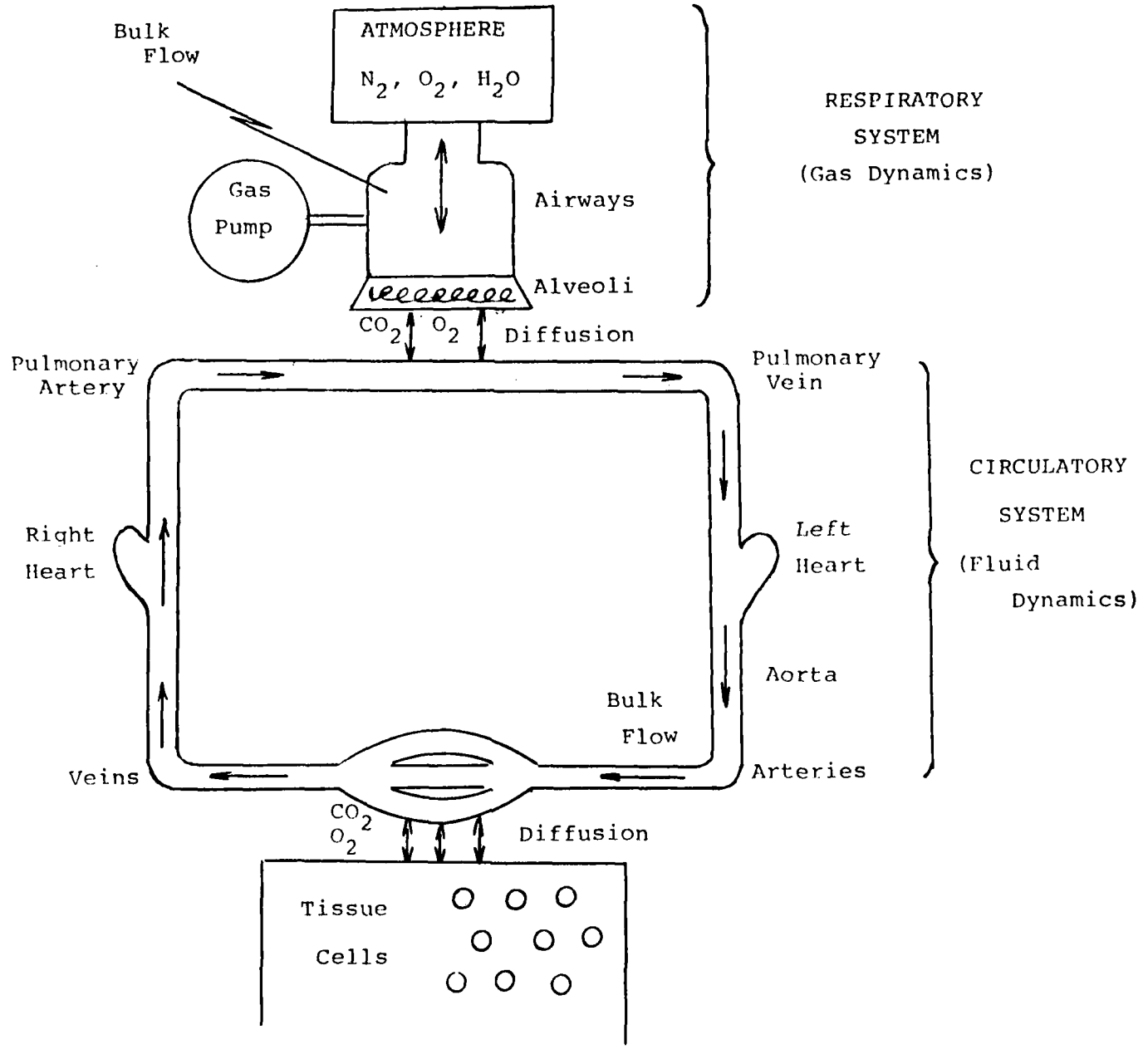

Bulk
Flow
ATMOSPHERE
N_2, O_2, H_2O
RESPIRATORY
SYSTEM
(Gas Dynamics)
Gas
Pump
Airways
Alveoli
CO_2
O_2
Diffusion
Pulmonary
Artery
Pulmonary
Vein
CIRCULATORY
SYSTEM
(Fluid
Dynamics)
Right
Heart
Left
Heart
Aorta
Bulk
Flow
Veins
Arteries
CO_2
O_2
Diffusion
Tissue
Cells

Figure 1

$j = O_2, CO_2, H_2O, N_2$ to other variables such as arterial concentrations $C_{a(j)}$, venous concentrations $C_{v(j)}$, inspired and expired volumes V_I, V_E, and control parameters consisting of blood flow rate $\dot{Q}_B$ and respiratory minute volume RMV. These latter variables are under CNS control via sensors for $C_{a(j)}$ in the carotid body and $C_{B(CO_2)}$ in the brain. Equilibrium equations relating $P_{A(j)}$ and $C_{a(j)}$ must also be written.

To illustrate how the transport delays are incorporated, we refer to the simplified schematic of Figure 2. From this, we see

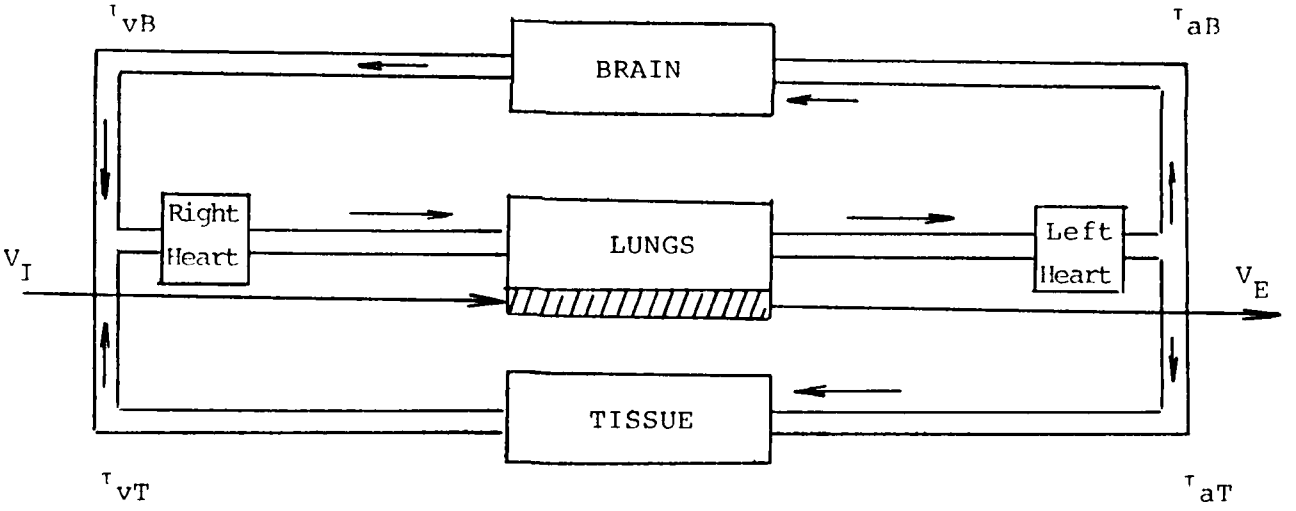


Figure 2

that transport delays τ come into play when relating the arterial concentrations $C_{aB(j)}(t)$, $C_{aT(j)}(t)$ at the brain and tissue entrance to the arterial concentrations $C_{a(j)}(t)$ at the lung exit. Delays are also appropriate in relating the mixed venous concentrations $C_{v(j)}$ entering the lungs with the venous concentrations $C_{vB(j)}$, $C_{vT(j)}$ at the brain and tissue exits. Using delay parameters as depicted in Figure 2, we have

$$C_{aB(j)}(t) = C_{a(j)}(t-\tau_{aB})$$
$$C_{aT(j)}(t) = C_{a(j)}(t-\tau_{aT})$$
$$C_{v(j)}(t) = g(C_{vB(j)}(t-\tau_{vB}), C_{vT(j)}(t-\tau_{vT})).$$

The delays themselves are in general functions of some of the variables (e.g., $\dot{Q}_B$ and thus $C_{a(O_2)}$ and $C_{a(CO_2)}$) in the models. However, often it is useful in employing the models in studies to select "apparent" values for the delays by fitting them to data. Delays can also arise in modeling the feedback control loops (see [34]) since the sensors for CO_2 are in the brain and carotid body while control is at the level of expired volume rate $\dot{V}_E$. That is, instead of RMV one might equivalently focus on the control $\dot{V}_E = h(P_{A(CO_2)}(t-\tau_1), P_{A(CO_2)}(t-\tau_2))$ where τ_1, τ_2 are transport times from the alveoli to the brain and carotid body respectively.

However the detailed model equations are formulated, it should be clear that one is lead naturally to feedback control

systems with delays

$$\dot{x}(t) = f(\alpha, x(t), x(t-\tau_1), \ldots, x(t-\tau_\nu)) \tag{2.1}$$

in which one often wishes to "identify" the parameters $q = (\alpha, \tau_1, \ldots, \tau_\nu)$.

Models for respiration in certain insects have also been investigated [11], [12] and these models involve a completely different type of system for which identification procedures are needed. Consider a tube (trachea - through which gas flows) of length L and radius r (cross-section $S = \pi r^2$) as depicted in Figure 3. Denote by e the thickness of the wall of the tube and by S_1 the cross-section of blood surrounding the tube.

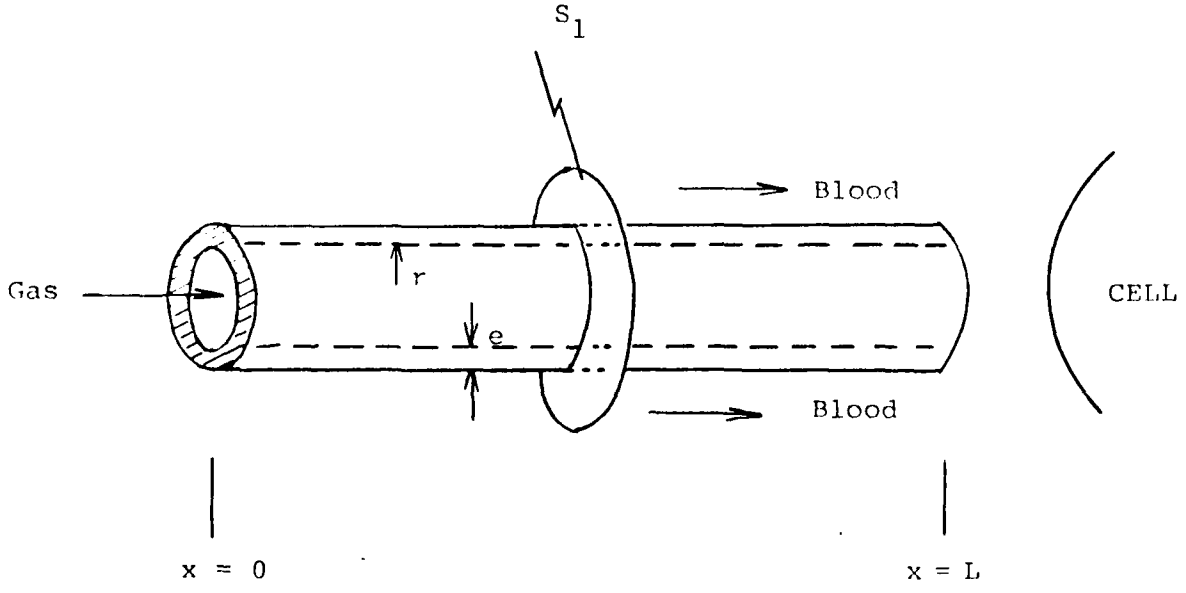


Figure 3

Using mass balance and considering (i) diffusion along the axis of the tube (radial diffusion assumed negligible), (ii) convection related to movement of the gas, and (iii) diffusion across the longitudinal wall, one can write equations which the partial pressures of oxygen, carbon dioxide, and nitrogen must satisfy. Letting y_1, y_2, y_3 denote the partial pressures in the tube of O_2, CO_2, N_2 respectively and w_1, w_2, w_3 denote the corresponding partial pressures in blood of these substances, one finds (the first, second, and third terms on the right in each equation represent axial diffusion, convection and diffusion across the wall respectively):

$$\frac{\partial y_j}{\partial t} = q_j \frac{\partial^2 y_j}{\partial x^2} - \frac{1}{PS} \frac{\partial}{\partial x} (V y_j) - D_j \frac{2}{re} (y_j - w_j),$$

(2.2)

$$\frac{\partial w_j}{\partial t} = \tilde{q}_j \frac{\partial^2 w_j}{\partial x^2} - \frac{\dot{Q}}{S_1} \sigma_j \frac{\partial w_j}{\partial x} - D_j \frac{2\pi r}{eS_1} (w_j - y_j),$$

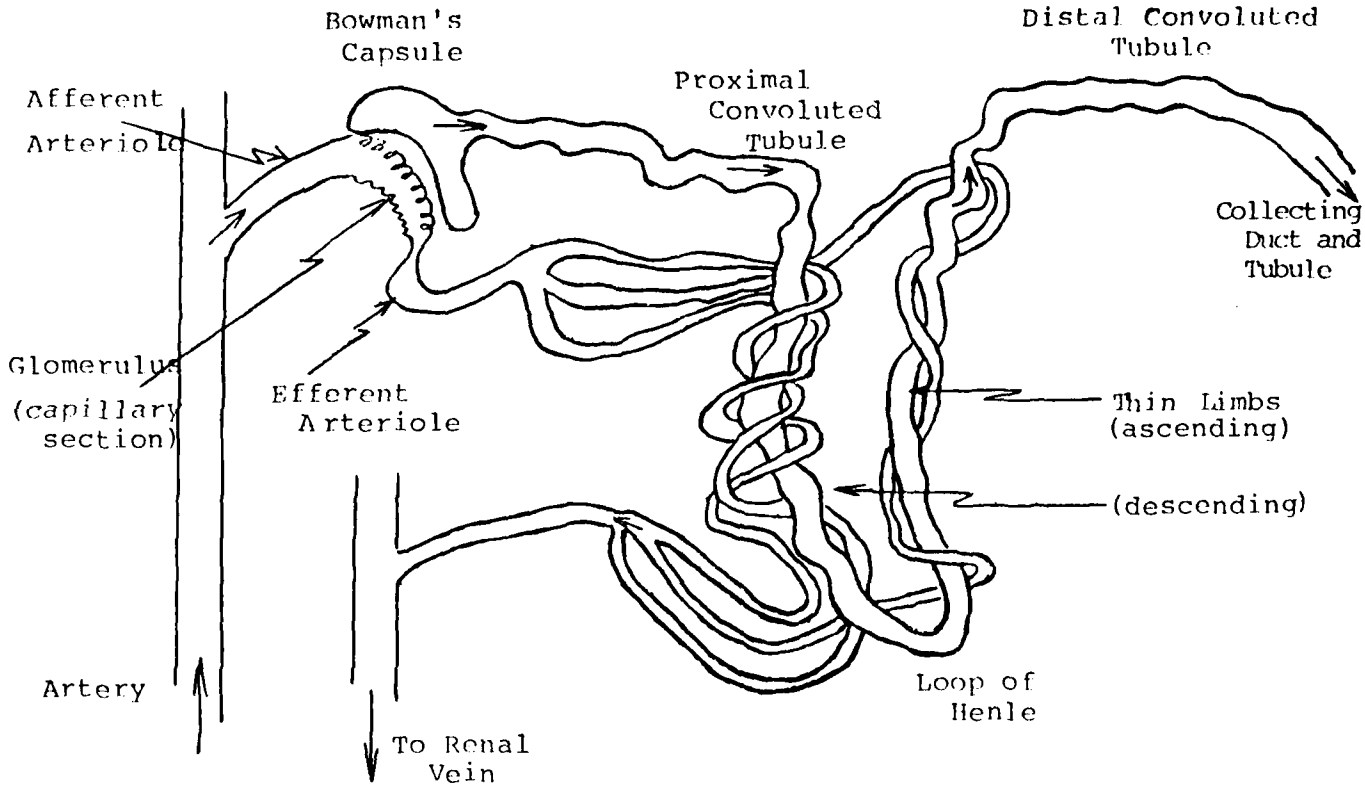
$$j = 1,2,3.$$

Here the parameters $q_j, \tilde{q}_j, D_j$ represent coefficients of diffusion, V is the velocity of the convective flow of the gas, P is atmospheric pressure, and the σ_j denote coefficients of solubility. Of course, $\dot{Q}$ is the blood flow rate. Among the important parameters to be estimated in use of such a model are the diffusion and solubility coefficients $q_j, \tilde{q}_j, D_j, \sigma_j$.

Systems of partial differential equations arising from convective-diffusive phenomena also are found in numerous models

for renal function. These models are usually based on mechanisms present in the nephron, the basic functional unit of the kidney (approximately one million per kidney) for which a schematic is given in Figure 4. Among the important mechanisms for transport between the blood and fluid in the nephron are filtration (transport from the plasma to glomerular fluid in Bowman's capsule), reabsorption (from the tubule back to the plasma; some of this involves active transport, some passive), and secretion (transport from plasma to urine which can be active or passive). For example, as fluid moves down the tubule, NaCl is removed actively, H_2O is removed passively from the tubule, leaving the concentration of urea higher in the tubule than in the perfusing plasma. Urea is then reabsorbed passively. The flow velocity in the tubule thus affects directly the amount of urine produced. For a high velocity flow only about 40% of the urea is reabsorbed whereas for low velocity flow (and hence a high rate of removal of NaCl and H_2O) up to 80% of the urea is reabsorbed.

Since approximately 7/8 of the total number of nephrons in a kidney are located in the cortex, much of the modeling in the literature (e.g., see [10],[30],[31],[32]) is focused on function in the renal cortex. Mathematical models are, not surprisingly, based on principles of conservation of mass and chemical species as well as the theory of non-equilibrium thermodynamics. A typical one-dimensional compartmental model might be derived employing a schematic as depicted in Figure 5. The basic equation for such


Bowman's Capsule
Distal Convoluted Tubule
Proximal Convoluted Tubule
Afferent Arteriole
Collecting Duct and Tubule
Glomerulus (capillary section)
Efferent Arteriole
Thin Limbs (ascending)
(descending)
Artery
Loop of Henle
To Renal Vein

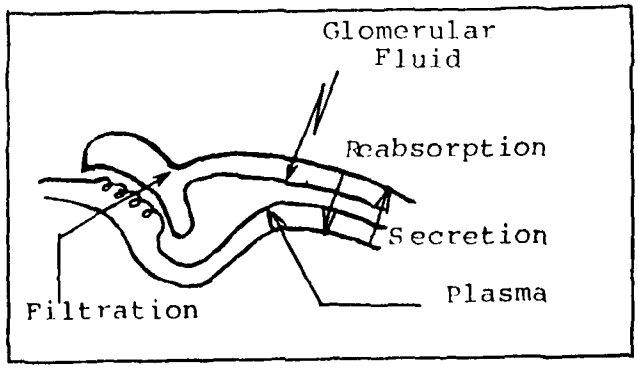

Glomerular Fluid
Reabsorption
Secretion
Plasma
Filtration

Figure 4

models is a one-dimensional convection-diffusion equation of the form

$$(2.3) \qquad \frac{\partial}{\partial t}(\rho_j A) = \frac{\partial}{\partial x}\left(q_j \frac{\partial \rho_j}{\partial x} A\right) + \frac{\partial}{\partial x}(V\rho_j) - J_j S$$

where ρ_j is the mass density of species j, A is the cross sectional area of the flow region, V is the volumetric flow rate in the axial direction x, $q_j \frac{\partial \rho_j}{\partial x}$ is the axial mass flux, S is the wall surface area per unit length and J_j is the mass flux of species j across the wall. We note that these equations result in a model of the same general form as that given in (2.2).

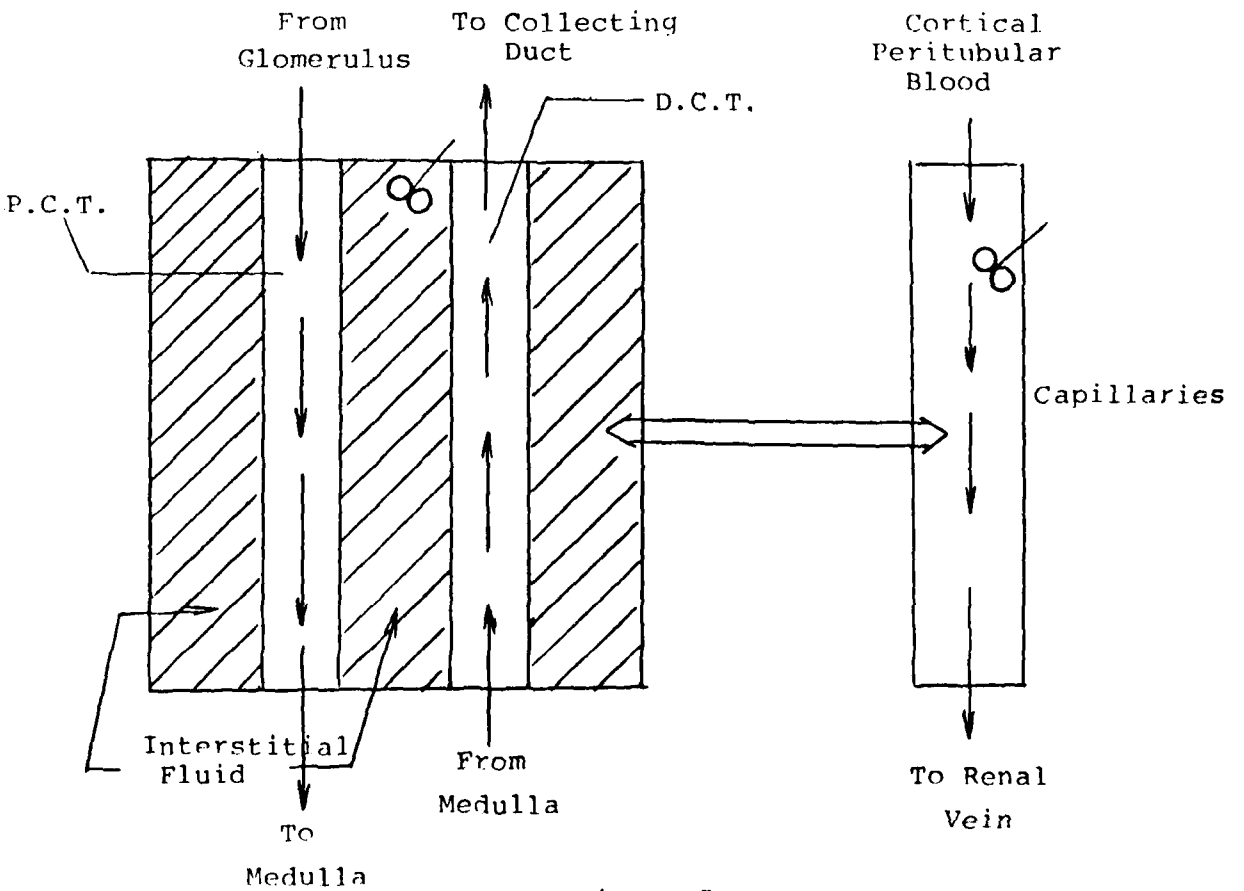


Figure 5

Summarizing, we see that models for respiratory and renal physiological control systems typically yield two basic types of systems for which parameter identification methods are required. One may have a delay system of the form (2.1) in which one seeks to estimate parameters $q = (\alpha, \tau_1, \ldots, \tau_\nu)$ including transport coefficients and the delays. Or one may be faced with identifying parameters including diffusion and solubility coefficients in parabolic distributed parameter (i.e., partial differential equation) systems such as (2.2) and (2.3). In both cases one must impose appropriate boundary and/or initial conditions and these also may contain unknown parameters for which estimates are needed.

These problems share certain inherent difficulties which we shall mention in section 4. Before doing so, however, it is necessary to review standard techniques that are readily available for parameter identification in ordinary differential equation models.

3. Parameter identification for ordinary differential equations

Consider a system which is modeled by an ordinary differential equation in R^n

$$\dot{x}(t) = f(q,t,x(t)), \quad 0 \le t \le T,$$
$$x(0) = x_0, \tag{3.1}$$

where q is a vector parameter in R^μ to be determined by observations of the system. A set of observations $\hat{y}_i \in R^\ell$, $i = 1,\ldots,m$, for $y(t_i) = Cx(t_i)$, $0 \le t_1 < \ldots < t_m \le T$ is given with C an $\ell \times n$ matrix. One wishes to use these to select a "best-fit" value $\bar{q}$ in Q; here Q is a given compact constraint set (admissible parameter values). A rather standard formulation is the least-squares fit-to-data in which one seeks to choose $q \in Q$ so as to minimize

$$J(q) = \frac{1}{2} \sum_{i=1}^{m} |y(t_i;q) - \hat{y}_i|^2 \tag{3.2}$$

where $y(t_i;q) = Cx(t_i;q)$ with x the solution of (3.1) corresponding to q. In practice, one turns to iterative methods to solve such problems.

Among the minimization techniques one might consider for these problems are those classified as direct search methods (for a detailed discussion of some typical methods in this class, see [10]). These methods consist of procedures which generate a sequence of trial solutions for minimizing J. Examination of a

given trial solution is by simple comparison, the result being used to indicate further steps in the search procedure. These methods, which make use of only functional evaluations, are attractive in some cases if one suspects that the function J to be minimized is not smooth, but they are slow and usually are quite inefficient when highly accurate solutions are desired. Furthermore, they have been developed heuristically and no proofs of convergence have been given. Indeed, while they might be quite useful in indicating the general location of a minimizing point, often their performance with regard to actual convergence is poor. One might expect that iterative methods which involve use of more of the information available about J (e.g., derivatives of various orders) would give superior performance in practical situations.

Among a number of methods of this latter type are some in a general class of methods, descent methods, which generate a sequence $\{q^k\}$ of approximations so that $J(q^{k+1}) \leq J(q^k)$. Recall that in many cases minimizing J is equivalent to seeking solutions of $J'(q) = 0$. If one applies Newton's method to this latter problem, one obtains the iterative procedure

$$q^{k+1} = q^k - [J''(q^k)]^{-1}J'(q^k). \tag{3.3}$$

We further recall that this is not a descent method, but under reasonable assumptions one can prove (see [29]) that there do exist

$\alpha_k > 0$ such that

$$q^{k+1} = q^k - \alpha_k [J''(q^k)]^{-1} J'(q^k) \tag{3.4}$$

is a descent method. This iterative formula defines the Damped Newton's method, which is a special case of the general iterative procedures defined by

$$q^{k+1} = q^k + \alpha_k p^k \tag{3.5}$$

where the "p^k" are called directions for the method. The analysis and development of many methods are concerned with how one is to choose the "directions" p^k and the "steplengths" α_k.

To motivate a choice of "good" directions p^k, consider a function J that is "quadratic" or "elliptic" in its behavior near a minimizing point $\bar{q}$ in R^2. (A similar argument can be carried out in R^n.) Then the level surfaces of J (actually their projection onto the R^2 plane, i.e., $\{q | J(q) = M\}$ for constants M) are given by closed curves as depicted in Figure 6.

At any point q on a level curve, the gradient vector $J'(q) = \nabla J(q) = (\frac{\partial J}{\partial q_1}, \frac{\partial J}{\partial q_2})$ is an outward normal vector. It is clear that for "good" directions to move toward $\bar{q}$, one wants to choose directions that are "downhill", i.e., directions given by vectors p satisfying $p \cdot J'(q) < 0$. Under reasonable assumptions

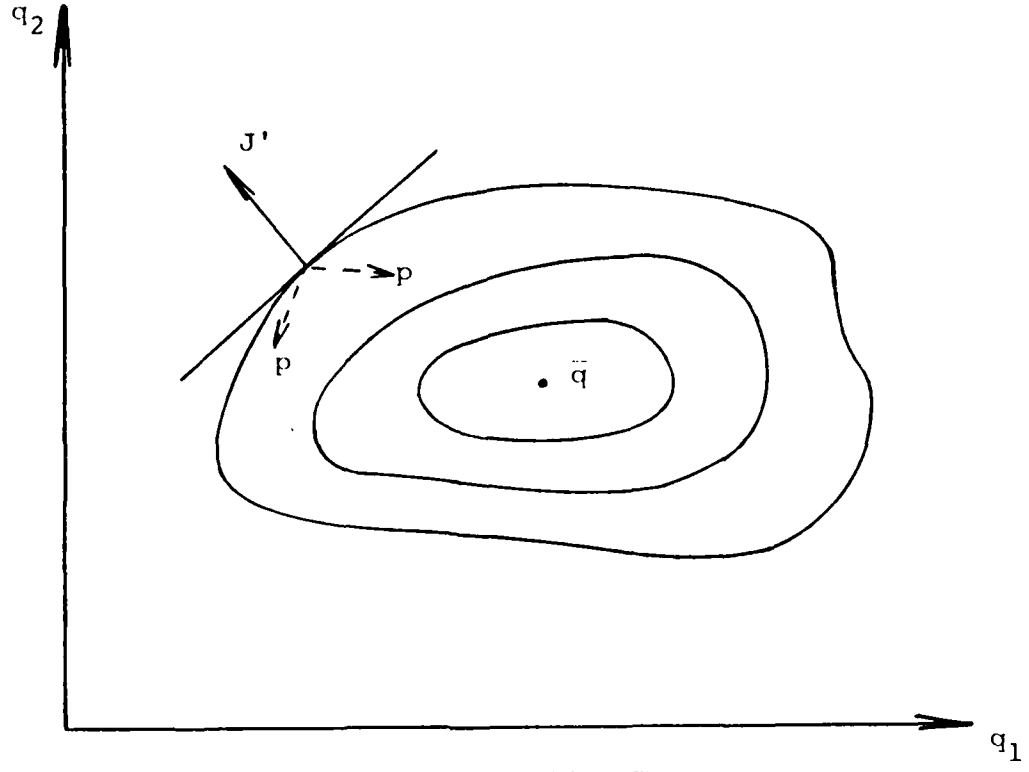


Figure 6

on J (see [21], [29]), one can show that if the p^k in (3.5) are chosen in a "downhill" manner, then it is possible to choose the steplengths $\{\alpha_k\}$ so that the resulting iterative procedure is a descent method and the iterates $\{q^k\}$ converge to a solution $\bar{q}$ of $J'(q) = 0$.

From the geometric considerations above, one can surmise that the "best" direction would be in the direction of -J' since clearly no other direction can give a larger local decrease in J. Ideas such as this are behind the steepest descent (or gradient) method where p^k is chosen as $p^k = -J'(q^k)$, which in the case

of (3.2) results in the iterative formula

$$q^{k+1} = q^k - \alpha_k J'(q^k) \tag{3.6}$$

where $J'(q^k) = \sum_{i=1}^{m} \frac{\partial y}{\partial q}(t_i;q^k)^T\{y(t_i;q^k) - \hat{y}_i\}$. (We assume here that one is using the usual Euclidean norm in R^μ. For "steepest descent" directions with respect to other norms in R^μ, see [29, p.245].) While (3.6) represents an intuitively appealing choice based on a "best" local strategy, the convergence properties of the resulting method are unfortunately much poorer than one might expect. One heuristic explanation sometimes offered for this poor performance is phrased as an "instability under small perturbations" and is depicted in Figure 7.

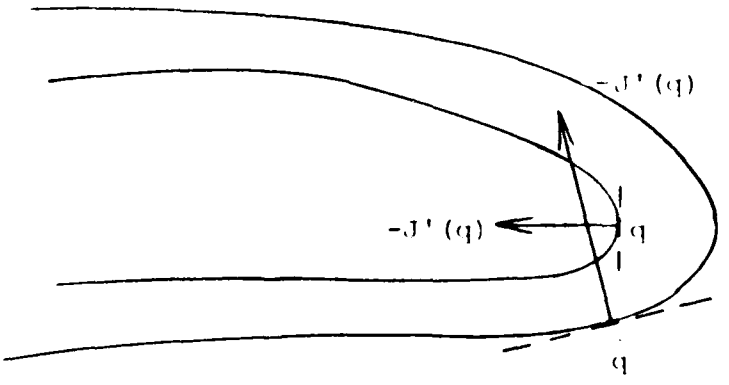


Figure 7

In this figure q represents the point given by precise computation while $\hat{q}$ is the actual computed value (different from q due to machine errors and the failure to compute exactly the previous descent step). From the figure it is clear that in extreme situations one could actually obtain a direction almost orthogonal to the one desired. In reality the explanation for the poor performance of the steepest descent method in the general case is somewhat more involved and is probably related to the fact that the steepest descent directions (when the method is applied to a quadratic function $J(q) = qAq$, $A > 0$) are in the limit asymptotic to just two directions [21]. Even though for "nice" J (e.g., $J(q) = qAq$ with $A > 0$) one can establish that convergence is geometric, i.e.,

$$J(q^{k+1}) \leq \frac{(\lambda_\mu - \lambda_1)^2}{(\lambda_\mu + \lambda_1)^2} J(q^k)$$

where $\lambda_1, \ldots, \lambda_\mu$ are eigenvalues of A, it is fair to say that the practical usefulness of the steepest descent method in locating precisely minimizing points for general functions J is probably overrated. (It is common to observe extremely slow convergence for this method in the vicinity of the minimizer.)

In a careful analysis of the difficulties with convergence of the steepest descent method, one actually discovers that the method essentially tries to approach a minimum in a two-dimensional subspace via use of a very limited choice of directions in the

iterative formula (3.6). One way to alleviate this problem is to ensure that one uses an adequate supply of directions by choosing the k^{th} direction p^k in (3.5) so that it is mutually orthogonal to p^{k-1}, $p^{k-2},\ldots,p^{k-(\mu-1)}$. This idea leads to the conjugate gradient or conjugate directions methods, the most useful general minimization methods currently available. The basic idea underlying these methods grew out of studies of minimization techniques for quadratic functionals $J(q) = \frac{1}{2}\, qAq + B\cdot q + E$, $A > 0$, on R^μ. These studies [29] reveal that if one chooses mutually A-orthogonal directions $p^1,\ldots,p^\mu$ (i.e., $p^iAp^j = \delta_{ij}$) and uses them in the iterative formula (3.5) with a proper choice of the $\{\alpha_k\}$ (obtained by minimization procedures - see the discussion below), the resulting sequence $\{q^k\}$ converges in at most μ steps to the unique minimizing point for J. (For a derivation of the conjugate direction methods based on Fourier expansion ideas which also yield the above results, see Chapter 10 of [23].) Before turning to a description of one popular generalization of these conjugate direction methods, we mention briefly that there are several possible ways for choosing the steplengths α_k in (3.5). Among so-called step-length algorithms probably the most often used are those based on the minimization principle

$$J(q^k+\alpha_k p^k) = \min\{J(q^k+\alpha p^k) \mid \alpha \in R^1\}. \tag{3.7}$$

That is, one chooses α_k so that J is minimized along the line $\{q|q = q^k + \alpha p^k, \alpha \in R^1\}$. This clearly always leads to a descent method. In complex problems one usually does not insist on precisely the minimizing α, but may use a one-dimensional search or some type of interpolation scheme to find an approximate to the minimizing α. For a discussion of other methods (Curry, Altman, majorization, Goldstein) in addition to (3.7) see [29, p.249-257].

One of the best known conjugate direction methods (the Fletcher-Reeves algorithm) uses the minimization principle for choice of step-lengths and chooses the direction "near" to those of steepest descent or gradient, i.e., those in (3.6), but modified to give conjugate directions (in the event J is quadratic). It is commonly referred to "the" conjugate-gradient method. (This is something of a misnomer since there are other algorithms - e.g., that of Daniel [29] - that are also called conjugate-gradient methods). Letting G^k denote the gradient of J at q^k, i.e., $G^k = J'(q^k)$, we define this iterative procedure by

$$p^0 = -G^0$$

$$p^{k+1} = -G^{k+1} + \beta_k p^k \tag{3.8}$$

$$\beta_k = |G^{k+1}|^2/|G^k|^2 \tag{3.9}$$

$$q^{k+1} = q^k + \alpha_k p^k \tag{3.10}$$

where α_k is chosen according to the minimization principle (3.7). We note that if $G^k = 0$, we have that q^k is a critical point of J and the algorithm is terminated. In the event that J is quadratic one can show that the directions (3.8) are conjugate directions. The method is obviously a descent method. In practice when applying this algorithm to problems with general J it is common to restart the method every μ^{th} step (or every $(\mu+1)^{st}$ step as suggested in [16]) by taking a purely steepest descent step. That is, one replaces (3.9) by

$$\beta_k = \begin{cases} 0 & \text{if } k+1 = M\mu \text{ for some positive integer } M \\ |G^{k+1}|^2/|G^k|^2 & \text{otherwise.} \end{cases}$$

For a discussion of the convergence properties of the Fletcher-Reeves algorithm, one should consult [29]. We note that some authors [21] erroneously assume that convergence of these algorithms in a finite number of steps for quadratic J automatically implies quadratic convergence in the general situation. For general J the methods appear not to be that well-behaved [29, p.512]. Other conjugate-direction type methods are also discussed in Chapter 8 of [29].

A class of popular quasi-Newton methods that are related to the above procedures is based on applying the Gauss-Newton scheme to a modified functional $\tilde{J}$. Given an estimate q^k, one linearizes

$y(t_i;q)$ in (3.2) about q^k obtaining the modified functional

$$(3.11)\quad \tilde{J}_k(q) = \frac{1}{2}\sum_{i=1}^{m} |y(t_i;q^k) + \frac{\partial y}{\partial q}(t_i;q^k)\{q-q^k\}-\hat{y}_i|^2.$$

One then chooses $q = q^{k+1}$ so as to minimize this functional or rather, to satisfy $\tilde{J}_k'(q^{k+1}) = 0$. This yields an equation for q^{k+1} given by

$$\sum_{i=1}^{m} \frac{\partial y}{\partial q}(t_i;q^k)^T \frac{\partial y}{\partial q}(t_i;q^k)\{q^{k+1}-q^k\} = -\sum_{i=1}^{m} \frac{\partial y}{\partial q}(t_i;q^k)^T\{y(t_i;q^k)-\hat{y}_i\}$$

or, using the notation $\partial_{ki} \equiv \frac{\partial y}{\partial q}(t_i;q^k)$,

$$(3.12)\quad q^{k+1} = q^k - \{\sum_{i=1}^{m} \partial_{ki}^T \partial_{ki}\}^{-1} \sum_{i=1}^{m} (\partial_{ki})^T\{y(t_i;q^k)-\hat{y}_i\}.$$

A modification of this algorithm (usually called the Levenberg-Marquardt algorithm) is used in a standard IMSL package that is widely available and involves the iterative formula

$$(3.13)\quad q^{k+1} = q^k - \{\lambda_k D_k + \sum_{i=1}^{m} \partial_{ki}^T \partial_{ki}\}^{-1} \sum_{i=1}^{m} (\partial_{ki})^T\{y(t_i;q^k) - \hat{y}_i\}$$

where $D_k = \text{diag}(\sum_{i=1}^{m} \partial_{ki}^T \partial_{ki})$ and λ_k is a parameter which can be chosen to insure that (3.13) is a descent method.

In all of the above procedures ((3.6), (3.8)-(3.10), (3.12), (3.13)), one needs $\frac{\partial y}{\partial q}$, which satisfies a linearized variational equation, as well as y, which satisfies the original nonlinear

equation (3.1) (or rather x does and $y = Cx$, $\frac{\partial y}{\partial q} = C\,\frac{\partial x}{\partial q}$). These must be computed at each iterative step, even though the iteration itself is in a finite-dimensional parameter space (recall $q \in Q \subset R^{\mu}$). Thus at each step of the algorithms, one must solve (3.1) for $t \to y(t;q^k) = Cx(t;q^k)$. For $\frac{\partial y}{\partial q}$ one can either solve for the fundamental matrix Φ_k of the linear variational equation

$$\dot{\Phi}_k(t) = \frac{\partial f}{\partial x}(q^k,t,x(t;q^k))\Phi_k(t)$$

and then employ a standard representation formula for $\frac{\partial x}{\partial q}$ in terms of Φ_k (see [6]), or, one can (this is most often done) use a difference approximation for $\frac{\partial y}{\partial q}$.

While the above methods lead to an iteration in a finite-dimensional parameter space, there is another widely publicized method, quasilinearization (see [6], and the references therein), which leads to a simultaneous iteration in a function space and parameter space. Given x^k and q^k, one lets $t \to x^{k+1}(t;q^{k+1})$ be the solution of

$$\dot{x}^{k+1}(t) = \frac{\partial f}{\partial x}(q^k,t,x^k(t))\{x^{k+1}(t) - x^k(t)\} +$$

$$\frac{\partial f}{\partial q}(q^k,t,x^k(t))\{q^{k+1}-q^k\} + f(q^k,t,x^k(t)) \tag{3.14}$$

$$x^{k+1}(0) = x_0,$$

where q^{k+1} is chosen so as to minimize

$$(3.15) \qquad \hat{J}(q^{k+1}) = \frac{1}{2} \sum_{i=1}^{m} |Cx^{k+1}(t_i;q^{k+1}) - \hat{y}_i|^2.$$

That is, we require that (3.14) and

$$(3.16) \qquad \frac{\partial}{\partial q^{k+1}} \frac{1}{2} \sum_{i=1}^{m} |Cx^{k+1}(t_i;q^{k+1}) - \hat{y}_i|^2 = 0$$

be satisfied in choosing x^{k+1}, q^{k+1}. The algorithm, which is explained in detain in [6], entails the following idea. Observe that, given x^k,q^k, the equation (3.14) is a linear variational equation for x^{k+1} in terms of q^{k+1}. One can thus use a representation formula for x^{k+1} in terms q^{k+1} and this can be substituted into (3.16) which is then solved for q^{k+1}.

Early advocates of the quasilinearization algorithm claimed that it converged quadratically (when it converged). If this were true, it might offer significant advantage over standard routines such as the Gauss-Newton (3.12) or (3.13) above which is usually quadratically convergent only when the model is <u>exact</u> (i.e., there exists a choice $\bar{q}$ so that $J(\bar{q}) = 0$ in (3.2)).

Both theoretical and numerical studies comparing the Gauss-Newton and quasilinearization algorithms have been carried out and the results reported in [1], [6], [18]. For the identification problems formulated here, it was found that from a theoretical viewpoint, the convergence properties of both the quasilinearization and Gauss-Newton algorithms are essentially equivalent (quadratic convergence when there is an exact fit of the model to the data;

at best linear convergence otherwise). In the numerical studies little difference between the algorithms was found in the number of iterations to convergence. The quasilinearization algorithm was faster per iteration (requiring fewer equivalent function evaluations than Gauss-Newton) but required much larger amounts of storage. The Gauss-Newton algorithm was usually more accurate and much less complex to program.

Results of these investigations appear to favor a simple Gauss-Newton algorithm applied directly to J over the quasilinearization algorithm for problems of the type formulated in this section.

A slightly different (but in some respects equivalent) formulation of the parameter identification problem for (3.1) seeks a value $\bar{q}$ that maximizes a "likelihood" function instead of a minimizer for J given in (3.2). That is, one seeks to choose $\bar{q} \in Q$ so as to maximize the <u>likelihood function</u>

$$L(q) = \sum_{i=1}^{m} \ell n\, g(\hat{y}_i - y(t_i;q)), \tag{3.17}$$

a heuristic foundation of which we shall explain briefly here. (A more detailed explanation of these procedures can be found in almost any standard text on estimation - e.g., see [14], [15], [22], [28].) A solution $\bar{q}$ of this problem is then called a <u>maximum likelihood estimator</u> (MLE) for q^*, the "true" parameter value (which, of course, may not exist). MLE algorithms are based on

the following considerations. The observations $\{\hat{y}_1,\ldots,\hat{y}_m\}$ are assumed to be corrupted by random measurement noises $\{N_1,\ldots,N_m\}$ so that we may write $\hat{y}_i = y(t_i;q^*) + N_i$, $i = 1,\ldots,m$. Assuming that $N_1,\ldots,N_m$ are independent random variables with identical probability density functions g, the joint density function is then given by $\tilde{g}(\eta_1,\eta_2,\ldots,\eta_m) = \prod_{i=1}^{m} g(\eta_i)$. Intuitively, the function $\tilde{g}$ should possess a maximum at those values of $(\eta_1,\ldots,\eta_m)$ that are most likely to occur. A procedure for estimating q^* might reasonably be devised on the basis that the observed values of $N_i = \hat{y}_i - y(t_i;q^*)$ correspond to those that are most likely to occur, that is, $\tilde{g}(N_1,\ldots,N_m) = \max \tilde{g}(\eta_1,\ldots,\eta_m)$. Defining the function $G(q) = \tilde{g}(\hat{y}_1-y(t_1;q),\ldots,\hat{y}_m-y(t_m;q))$ we therefore might seek a value $\bar{q}$ that yields a maximum for G. For technical reasons, it is more convenient to maximize the natural log of this function so that one defines the likelihood function

$$L(q) = \ell n\ G(q) = \sum_{i=1}^{m} \ell n\ g(\hat{y}_i - y(t_i;q))$$

and equivalently seeks a maximizer $\bar{q}$ (called an MLE for q^*) of L. In certain cases (see [15]) maximizing L is completely equivalent to minimizing the least squares criterion J defined by (3.2). More generally, the MLE procedures reduce to an algorithm to determine a solution $\bar{q}$ of $L'(q) = 0$.

4. Inherent difficulties in FDE and PDE identification

In all of the techniques outlined in section 3 for ordinary differential equations, we noted that at each iterative step one usually required solution of the original system (3.1) as well as perhaps solution of an associated variational equation. The system (3.1) and any linearized variational equation involves systems in finite-dimentional state space R^n. For delay systems and distributed parameter systems such as those discussed in connection with the physiology models of section 2, one must deal with infinite dimensional "state" processes. If applied directly to such systems the iterative methods of section 3 would entail huge storage requirements and would, in many cases of practical interest, prove unwieldly and indeed unfeasible.

Even if one had unlimited storage capabilities so that the above mentioned difficulties might effectively be circumvented, there are even more serious problems inherent in using the methods of section 3 directly for certain identification problems involving delay systems. As we saw in section 2, it is often very important to estimate, among other parameters, the delays in a functional differential equation model. A careful inspection of the techniques outlined in section 3 reveals that one must deal with the derivatives $\frac{\partial x}{\partial q}$ or, in the case one is identifying the delays in $\dot{x}(t) = f(\alpha, x(t), x(t-\tau_1), \ldots, x(t-\tau_\nu))$, one must make use of $\frac{\partial x}{\partial \tau_i}$ as well as $\frac{\partial x}{\partial \alpha}$. Even if f is extremely smooth, these derivatives need not exist. Consider, for example,

$$\dot{x}(t) = x(t-\tau), \quad t > 0$$

(4.1)

$$x(\theta) = \phi(\theta) \quad , \quad \theta \leq 0$$

where the initial function ϕ is defined by $\phi(\theta) = \frac{1}{2}$, $-1 < \theta \leq 0$, $\phi(\theta) = 1$ for $\theta \leq -1$. For $\tau \leq 1$ this equation has solution $x(t) = \frac{1}{2} + \frac{1}{2}t$ on $[0,\tau]$. For $\tau > 1$, say $\tau = 1 + \varepsilon$, we find the solution is given by $x(t) = \frac{1}{2} + t$ on $[0,\varepsilon]$, $x(t) = \frac{1}{2} + \varepsilon + \frac{1}{2}(t-\varepsilon)$ on $(\varepsilon,1+\varepsilon]$. If one then fixes t_i in $(0,1)$, it is quite easy to demonstrate that $\tau \to x(t_i;\tau)$ is not differentiable at $\tau = 1$; i.e., $\frac{\partial x}{\partial \tau}(t_i;1)$ does not exist. Hence one cannot apply directly the methods outlined in section 3 for identification of delay parameters in examples such as this one. The methods we describe in the next section can be used to overcome such difficulties as well as those inherent in the infinite-dimensionality of the state in such systems.

5. Techniques for FDE and PDE identification

We turn finally to methods for parameter estimation in problems involving functional differential or partial differential equations. We shall for ease in exposition assume a least-squares formulation of the basic identification problem. Thus we seek to minimize J over Q with

$$J(q) = \frac{1}{2} \sum_{i=1}^{m} |y(t_i;q)-\hat{y}_i|^2 \tag{5.1}$$

as in section 3 except now y is the output for either an FDE or a PDE. For example, we might have $y(t) = Cx(t)$ with x the solution of an underlying delay system such as (2.1) or $y(t) = \text{col}\{Cu(t,x_1),\ldots,Cu(t,x_p)\}$ with u the solution of a system of partial differential equations. Before turning to specific cases which differ slightly in detail, we outline briefly the fundamental ideas involved in development of the methods we shall discuss.

We first rewrite the system (FDE or PDE) as an abstract equation in an appropriately chosen Hilbert space Z:

$$\begin{aligned} \dot{z}(t) &= \mathscr{A}(q)z(t) + G(t) \\ z(0) &= z_0. \end{aligned} \tag{5.2}$$

Here $\mathscr{A}$ may be either a linear or nonlinear operator depending on parameters $q \in Q$. The identification problem is reformulated in a corresponding manner so that one seeks $\bar{q} \in Q$ that minimizes

$$J(q) = \frac{1}{2} \sum_{i=1}^{m} |\Gamma(z(t_i;q)) - \hat{y}_i|^2 \tag{5.3}$$

where $y(t) = \Gamma(z(t_i;q))$ is an appropriately defined output.

The problem now has the appearance of a parameter identification problem for an ordinary differential equation except, of course, we are working in an infinite dimensional state space Z instead of R^n as in section 3. We attempt to reduce the problem for (5.2), (5.3) to an approximate one in a finite dimensional space. The approximation idea we employ is a classical one commonly referred to as the Ritz-Galerkin technique. We choose subspaces Z^N of Z with projections $P^N: Z \to Z^N$ and solve a least-squares problem for the system as it is "projected" or approximated in these subspaces. That is, we seek to minimize

$$J^N(q) = \frac{1}{2} \sum_{i=1}^{m} |\Gamma(z^N(t_i;q)) - \hat{y}_i|^2 \tag{5.4}$$

over Q subject to the approximate system

$$\begin{aligned} \dot{z}^N(t) &= \mathscr{A}^N(q) z^N(t) + P^N G(t) \\ z^N(0) &= P^N z_0. \end{aligned} \tag{5.5}$$

Our hope, of course, is that by clever choices of Z^N, P^N and $\mathscr{A}^N$, we might be able to insure that $z^N(t) \to z(t)$ and that solutions $\bar{q}^N$ of minimizing (5.4) subject to (5.5) will, as $N \to \infty$, approach some $\bar{q}$ in Q that is a solution of the original problem for (5.1).

We describe two particular choices for such schemes: In the case of FDE's we outline developments for spline subspaces Z^N while for PDE's we report on results involving modal (eigenfunction) subspaces Z^N.

FDE and spline approximations

We consider the nonlinear FDE

$$\dot{x}(t) = f(\alpha, x(t), x_t, x(t-\tau_1), \ldots, x(t-\tau_\nu)) + g(t)$$
$$x_0 = \phi \tag{5.6}$$

where x_t is the usual notation for the function $\theta \to x_t(\theta) = x(t+\theta)$, $-r = \tau_\nu \le \theta \le 0$, representing a functional dependence on the past in our system. In general we seek to determine from data (observations on Cx) values for the parameters $q = (\alpha, \tau_1, \ldots, \tau_\nu)$. By choosing $Z = R^n \times L_2(-r,0;R^n)$ with an appropriate inner product $\langle \cdot, \cdot \rangle$ and employing as state $z(t) = (x(t), x_t)$, where x is the solution of (5.6), we can rewrite (5.6) as an equivalent abstract system (5.2) with $\mathcal{A}(q)z = (f(\alpha, \psi(0), \psi, \psi(-\tau_1), \ldots, \psi(-\tau_\nu)), \psi')$ for $z = (\psi(0), \psi)$ in a judiciously chosen domain in Z (for details see [3]) and $z_0 = (\phi(0), \phi)$.

In describing the spline approximations we shall for ease in exposition restrict our considerations to first-order (piecewise linear) spline approximations for a scalar equation of the form (5.6). A general theory for arbitrary-order spline approxima-

tions is given in [7] for linear FDE and in [3], [20] for nonlinear FDE. We shall also assume (again only for ease in exposition) that the delays $\tau_1,\ldots,\tau_\nu$ and initial data ϕ are known. In this case, the theory for nonlinear system parameter estimation via spline approximations is given in [3], while the general nonlinear theory for estimation of parameters including delays and initial data (which is technically more difficult to describe) is given in [5].

Let e_j^N, $j = 0,1,\ldots N$, be the classical first-order spline functions on $[-r,0]$ satisfying $e_i^N(t_j^N) = \delta_{ij}$, where δ_{ij} is the Kronecker symbol and $t_j^N = -\frac{jr}{N}$. That is, we partition $[-r,0]$ into subintervals $[t_j^N, t_{j-1}^N]$ and define e_j^N to be 1 at t_j^N, 0 at t_{j+1}^N and t_{j-1}^N, and linear on $[t_{j+1}^N,t_j^N]$ and $[t_j^N,t_{j-1}^N]$ with e_j^N vanishing outside $[t_{j+1}^N,t_{j-1}^N]$. Let $Z^N = \text{span}\{\beta_0^N,\ldots,\beta_N^N\}$ where $\beta_j^N = (e_j^N(0),e_j^N) \in Z$. Then Z^N is an $N+1$ dimensional subspace of Z and for any $z^N(t) \in Z^N$, we can write $z^N(t) = \sum_{j=0}^N w_j^N(t)\beta_j^N$ where w_j^N are the generalized Fourier coefficients or "coordinates" of z^N relative to $\{\beta_0^N,\ldots,\beta_N^N\}$. Let P^N be the orthogonal projection of Z onto Z^N which is characterized by the condition

$$\langle P^N z - z,\ \beta_j^N\rangle = 0,\quad j = 0,1,\ldots,N,\quad z \in Z. \tag{5.7}$$

If $P^N z = \sum_{k=0}^N \alpha_k\beta_k^N$ then (5.7) is equivalent to

$$K^N\alpha^N = \text{col}(\langle z,\beta_0^N\rangle,\ldots,\langle z,\beta_N^N\rangle) \equiv h^N(z) \tag{5.8}$$

where $\alpha^N = \mathrm{col}(\alpha_0, \alpha_1, \ldots, \alpha_N)$ and K^N is the $(N+1) \times (N+1)$ matrix with elements $\langle \beta_i^N, \beta_j^N \rangle$. For first-order splines this is a tridiagonal matrix and equation (5.8) is easily solved for α^N (for higher order splines the analogue of K^N is a banded matrix so that computing P^N is also easily done in those cases).

Next we define $\mathscr{A}^N(q) = P^N \mathscr{A}(q) P^N$ which is readily computed since we know $\mathscr{A}$ and know how to compute P^N. The approximating equation (5.5) for (5.6) thus reduces to an ordinary differential equation in $w^N = \mathrm{col}(w_0^N, w_1^N, \ldots, w_N^N)$, the coordinates of z^N. One obtains

$$\dot{w}^N(t) = A^N(q) w^N(t) + (K^N)^{-1} \mathrm{col}(g(t), 0, \ldots, 0)$$

(5.9)

$$w^N(0) = (K^N)^{-1} h^N((\phi(0), \phi))$$

where $A^N(q)$ is the (in general nonlinear) operator

$$A^N(q) w^N(t) = (K^N)^{-1} h^N\left(\mathscr{A}(q)\left(\sum_{j=0}^{N} w_j^N(t) \beta_j^N\right)\right).$$

The problem of minimizing (5.4) subject to (5.5) thus reduces to an easily implemented problem for an ordinary differential equation (5.9) for which the standard techniques discussed in section 3 can readily be employed.

We have tested these ideas on a number of examples and present now a small sample of numerical results to illustrate our findings. In each case an IMSL package employing the Levenberg-

Marquardt algorithm (see (3.13)) was used to solve the ordinary differential equation approximating identification problems.

Example 5.1.

We return to the example of section 4 (see the comments following (4.1))

$$\dot{x}(t) = x(t-\tau), \quad t > 0,$$
$$x(\theta) = \phi(\theta),$$

with $\phi(\theta) = \frac{1}{2}$ for $-1 < \theta \leq 0$, $\phi(\theta) = 1$ for $\theta \leq -1$. We seek to identify a "true" value $\bar{\tau} = 1.0$ for the delay parameter even though $\frac{\partial x}{\partial \tau}(t_i;1)$ does not exist. We present the numerical findings in a tabular form with $\bar{\tau}^N$ the "converged" value (from the iterative IMSL package) of τ for the index of approximation N. Start-up values of $\tau^{N,0} = .9$ were used in each case below. Very similar results were obtained with start-ups $\tau^{N,0} = .2$ and 3.0. The true analytical solution was used as "data" on [0,1].

N	$\bar{\tau}^N$	No. of IMSL Iterates
2	.8031	3
4	.9080	3
8	.9686	5
16	.9950	6
32	1.0010	7
True Value	1.0000	

Example 5.2.

Consider the multiple delay example

$$\dot{x}(t) = 2x(t) + \alpha x(t-\tau) + x(t-2), \qquad t > 0,$$
$$x(\theta) = 1, \qquad -2 \leq \theta \leq 0,$$

for which "data" corresponding to "true" values $\bar{\alpha} = 3.0$, $\bar{\tau} = 1.0$ of the parameters $q = (\alpha,\tau)$ are easily generated on $[0,3]$. In the table below start-up values $\alpha^{N,0} = 2.6$, $\tau^{N,0} = 1.3$ were used for each N. For each N the IMSL package produced a "converged" estimate in 10 iterations.

N	$\bar{\alpha}^N$	$\bar{\tau}^N$
2	3.888	.9472
4	3.243	.9851
8	3.062	.9961
16	3.0159	.9991
32	3.0040	.9998
True Values	3.0	1.0

Example 5.3.

Finally we consider the nonlinear equation

$$\dot{x}(t) = 2x(t) + 5x(t-\tau) + \frac{\alpha x(t-2)}{1+x(t-2)}, \qquad t > 0,$$
$$x(\theta) = 1$$

in which we seek to estimate $q = (\alpha,\tau)$. An independent numerical method can be used to generate "data" on [0,4] corresponding to "true" values $\bar{q} = (\bar{\alpha},\bar{\tau}) = (3.0,1.0)$. Again the approximation methods described above worked quite well when applied to this example. For $N = 40$, the IMSL package converged in 21 iterations to values $\bar{q}^{40} = (\bar{\alpha}^{40},\bar{\tau}^{40}) = (3.0058,.995)$ from start-up values $q^{N,0} = (1.0,.5)$.

We conclude our brief discussion of the spline methods for FDE parameter identification with several remarks. First, higher order spline methods are easily implemented in the manner outlined above. Computational experiments by F. Kappel and colleagues in Graz reveal that (at least in the case of cubic splines) the resulting banded matrices pose no conceptual or practical difficulties in implementation. Vector systems are also easily treated using these approximations (see, for example, the column reactor identification example where $n = 8$ in [2]). Proofs of convergence for the fundamental approximations in the case of linear systems are given via use of abstract approximation theorems from semigroup theory (a Trotter-Kato approximation theorem) in [7] where error estimates for the "state" convergence $(z^N \to z)$ are also given. The fundamental theory for nonlinear systems is given in [3], [20]. As one might expect, one obtains that the piecewise-linear elements yield $O(\frac{1}{N})$ estimates of convergence, cubic elements converge like $O(\frac{1}{N^3})$, etc. We have made numerous computational tests (see [3], [4], [7], [20]) with these methods involving state approximation

only, parameter identification problems, and optimal control problems. Our experience has been similar to that realized in the use of finite-element methods for certain boundary value problems in that we observe convergence rates that are better than predicted by the theory. For example, the piecewise linear elements generate schemes that appear basically second order in behavior (i.e., errors like $O(\frac{1}{N^2})$) and this is observed not only in convergence of the states $z^N \to z$ but also in parameter estimates and in optimal controls and performance criteria.

PDE and modal approximations

To facilitate our discussion of modal approximation schemes we consider simple scalar nonlinear parabolic equations

$$\frac{\partial u}{\partial t} = \frac{q_1}{k}\frac{\partial}{\partial x}(p\frac{\partial u}{\partial x}) + q_2 u + f(q_4,t,x,u), \quad 0 < x < 1, \ t > 0, \tag{5.10}$$

with initial condition

$$u(0,x) = q_3\phi(x), \tag{5.11}$$

and boundary conditions

$$u(t,0) = u(t,1) = 0, \tag{5.12}$$

where p and k satisfy the usual Sturm-Liouville conditions

$(p,p',k$ continuous with $p > 0$, $k > 0)$. A general approximation framework (again based on Trotter-Kato type approximation theorems from semigroup theory) that includes as special cases both nonlinear parabolic and hyperbolic vector systems (with quite general boundary conditions allowed) is detailed in some of our joint efforts with K. Kunisch [8], [9]. We shall here only briefly indicate how one chooses the spaces and operators in the case of (5.10)-(5.12) so that we are in a special case of the general Ritz-Galerkin ideas outlined above.

To rewrite (5.10)-(5.12) as an abstract Cauchy problem we choose $Z = L_2(0,1)$ with inner product $\langle \phi,\psi \rangle = \int_0^1 \phi(x)\psi(x)k(x)dx$ and define (on an appropriately chosen domain in Z) the operator $\mathscr{A}(q)$ by $\mathscr{A}(q)\psi = \frac{q_1}{k}(p\psi')' + q_2\psi$. Then with a carefully defined nonlinear map F (see [8]) we find that (5.10)-(5.12) is equivalent in some sense (one which is sufficient for our purposes) to the equation in Z

$$\dot{z}(t) = \mathscr{A}(q)z + F(q,t,z(t)), \quad t > 0$$
$$z(0) = q_3\phi. \tag{5.13}$$

Sturm-Liouville operator theory can be employed to obtain a complete orthonormal set of eigenfunctions $\{\psi_j\}_{j=1}^{\infty}$ for $\mathscr{A}(\tilde{q})$ with $\tilde{q} = (1,0,\ldots,0)$ and these are used to define modal approximation subspaces $Z^N = \text{span}\,\{\psi_1,\ldots,\psi_N\}$. One then defines P^N and

$\mathcal{A}^N(q)$ in a manner similar to that for the FDE approximations discussed above; i.e., $\mathcal{A}^N(q) = P^N \mathcal{A}(q) P^N$ where P^N is the orthogonal projection $P^N z = \sum_{j=1}^{N} \langle z, \psi_j \rangle \psi_j$ of Z onto Z^N. An approximating ordinary differential equation ((5.5) with G replaced by $F(q,t,z^N(t))$) is then used to formulate and solve corresponding approximate identification problems in a (by-now) obvious manner.

We have tested the methods proposed here for both parabolic and hyperbolic examples and present a sample of our preliminary numerical findings for parabolic equations. "Data" for the examples below were generated by employing an independent method (Crank-Nicolson) to solve numerically the equations with the parameters q set equal to their "true" values $\bar{q}$.

Example 5.4

We consider the equation

$$u_t = .1\, u_{xx} + q_2 u$$

with boundary conditions (5.12) and initial condition (5.11) where $q_3 = 1$ and ϕ is taken as the piecewise linear continuous ("roof") function that is linear on $[0,\frac{1}{2}]$ and $[\frac{1}{2},1]$ and satisfies $\phi(0) = \phi(1) = 0$, $\phi(\frac{1}{2}) = 1$. A true value of $\bar{q}_2 = .8$ was chosen. For $N = 4$, a start-up value of $q_2^{4,0} = .25$ was selected and the IMSL package for the approximating problem produced a converged value $\bar{q}_2^4 = .8001$.

We considered then the equation

$$u_t = q_1 u_{xx} + .2u$$

with the same initial and boundary conditions. A true value $\bar{q}_1 = .1$ was taken. For $N = 4$ and start-up value $q_1^{4,0} = .25$, the estimate $\bar{q}_1^4 = .0999$ was obtained.

Example 5.5.

The nonlinear equation

$$u_t = .1u_{xx} + q_4 \frac{2}{1+u}$$

with boundary conditions (5.12) and initial condition (5.11) where ϕ is as in Example 5.4 (except $\phi(\frac{1}{2}) = 4$) was investigated with "true" parameter values $(\bar{q}_3, \bar{q}_4) = (5.0, 2.0)$. With start-up values $(q_3^{N,0}, q_4^{N,0}) = (1.0, 0.0)$ the following estimates were obtained.

N	$\bar{q}_3^N$	$\bar{q}_4^N$
4	5.2274	1.9254
8	5.1374	1.9741
16	5.0668	1.9845
True Values	5.0	2.0

The theory for modal approximations in the case of hyperbolic equations is only slightly more involved than that sketched above for parabolic systems. For example, an equation of the form

$$u_{tt} = q_1 u_{xx} + q_2 u_t + q_3 u + f(q_4,t,x,u)$$

can be written as a first order vector system employing appropriate Sobolev spaces (e.g., $Z = H_0^1 \times L_2$)

$$\frac{d}{dt}\begin{bmatrix} u \\ u_t \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ q_1\Delta + q_3 & q_2 \end{bmatrix}\begin{bmatrix} u \\ u_t \end{bmatrix} + \begin{bmatrix} 0 \\ f \end{bmatrix}$$

and a convergence analysis carried out with the aid of general spectral theorems and semigroup approximation results. For details as well as numerical examples, see [8],[9].

Acknowledgement: Two of my students, Patti Daniel and Jim Crowley, have contributed significantly to the numerical studies outlined in section 5 and I would like to express my deep appreciation for their continuing efforts on these and related investigations.

REFERENCES

[1] H.T.Banks, On identification problems arising in biomedical modeling, Séminaires IRIA: analyse et contrôle de systemes, 1972, 33-37.

[2] H.T.Banks, Computational difficulties in the identification and optimization of control systems, Proc. Workshop on Applied Control Theory to Renewable Resource Management and Ecology, Univ. Canterbury, Christchurch, New Zealand, Jan. 7-11, 1980, Springer-Verlag, to appear.

[3] H.T.Banks, Identification of nonlinear delay systems using spline methods, Proc. International Conference on Nonlinear Phenomena in the Math Sciences, Univ. Texas, Arlington, June 16-20, 1980, Academic Press, to appear.

[4] H.T.Banks, J.A.Burns, and E.M.Cliff, A comparison of numerical methods for identification and optimization problems involving control systems with delays, LCDS Tech. Rep. 79-7, November 1979, Brown University, Prov. R.I.

[5] H.T.Banks and P.Daniel, Estimation of delays and other parameters in nonlinear functional differential equations, to appear.

[6] H.T.Banks and G.M. Groome, Convergence theorems for parameter estimation by quasilinearization, J. Math. Anal. Appl. 42 (1973), 91-109.

[7] H.T.Banks and F.Kappel, Spline approximations for functional differential equations, J. Differential Eqns. 34(1979), 496-522.

[8] H.T.Banks and K. Kunisch, Parameter estimation techniques for nonlinear distributed parameter systems, Proc. International Conference on Nonlinear Phenomena in the Math. Sciences, Univ. Texas, Arlington, June 16-20, 1980, Academic Press, to appear.

[9] H.T.Banks and K. Kunisch, An approximation theory for nonlinear partial differential equations with applications to identification and control, to appear.

[10] H.T. Banks and P.J. Palatt, Mathematical Modeling in the Biological Sciences, LCDS Lecture Notes 75-1, June 1975, Brown University, Providence, R. I.

[11] J.Brocas and Y.Cherruault, Association des systèmes trachéens et circulatoires dans le transport des gaz respiratoires chez l'insecte aérien, International J. Bio-Med. Computing 4(1973), 173-204.

[12] J.Brocas and Y. Cherruault, Etude sur modèle mathématique des echanges gazeux alvéolo-capillaires - Rôle joué dans ces echanges par le presence d'un gaz inerte, Math. Biosci. 18 (1973), 313-333.

[13] J.Comroe, Physiology of Respiration, Year Book Med. Publ., Chicago, 1965.

[14] H.Cramér, Mathematical Methods of Statistics, Princeton Univ. Press, Princeton, N.J. 1946.

[15] R.Deutsch, Estimation Theory, Prentice-Hall, Englewood Cliffs, N.J. 1965.

[16] R.Fletcher and C.M.Reeves, Function minimization by conjugate gradients, Computer J. 7(1964), 149-154.

[17] F.Grodins, J.Buell, and A.Bart, Mathematical analysis and digital simulation of the respiratory control system, J. Applied Physiol. 22(1967), 260-276.

[18] G.M.Groome, Identification of Dynamical Systems, Ph.D. Thesis, Brown University, June, 1972.

[19] A.C.Jackson and H.T. Milhorn, Digital computer simulation of respiratory mechanics, Comp. Biomed. Res. 6(1973), 27-56.

[20] F.Kappel, Spline approximation for autonomous nonlinear functional differential equations, preprint, June 1980, to appear.

[21] J.Kowalik and M.R.Osborne, Methods for Unconstrained Optimization Problems, American Elsevier, New York, 1968.

[22] B.W.Lindgren, Statistical Theory, Macmillan, New York, 1962.

[23] D.G.Luenberger, Optimization by Vector Space Methods, John Wiley and Sons, New York, 1969

[24] M.C.Mackey and L. Glass, Oscillation and chaos in physiological control systems, Science 197(July, 1977), 287-289.

[25] H.T. Milhorn, The Application of Control Theory to Physiological Systems, Saunders, Philadelphia, 1966, Chapter 15.

[26] H.T.Milhorn, R.Benton, R.Ross, and A.C.Guyton, A mathematical model of the human respiratory control system, Biophysical J. 5(1965), 27-46.

[27] H.T.Milhorn and A.C.Guyton, An analog computer analysis of Cheyne-Stokes breathing, J. Appl. Physiol. 20(1965), 328-333.

[28] A.M.Mood and F.A.Graybill, Introduction to the Theory of Statistics, McGraw-Hill, New York, 1963.

[29] J.M.Ortega and W.C.Rheinboldt, Iterative Solution of Non-linear Equations in Several Variables, Academic Press, New York, 1970.

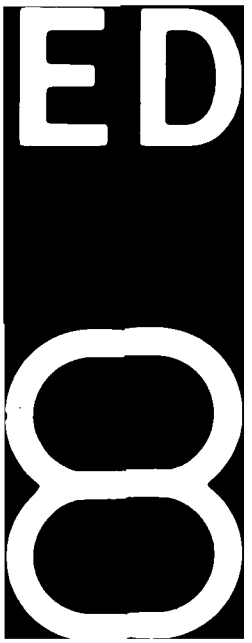
[30] P.J.Palatt, H.Sackin and R.I.Tanner, A hydrodynamic model of a permeable tubule, J. Theor. Biol. 44(1974), 287.

[31] P.J.Palatt and G.M.Saidel, A note on a simple counter-current exchange system, T.-I.-T. J. Life Sci. 1(1971), 123.

[32] P.J.Palatt, G.M.Saidel and J.Macklin, Transport processes in the renal cortex, J. Theor. Biol. 29(1970), 251-274.

[33] E. Selkurt, Physiology, Little, Brown and Co., Boston, 1966.

[34] D.M.Wiberg, J.W.Bellville, O.Brovko, R.Maine, and T.C.Tai, Modeling and parameter identification of the human respiratory system, IEEE Trans. Auto. Control, AC-24 (1979), 716-720.


ED
8